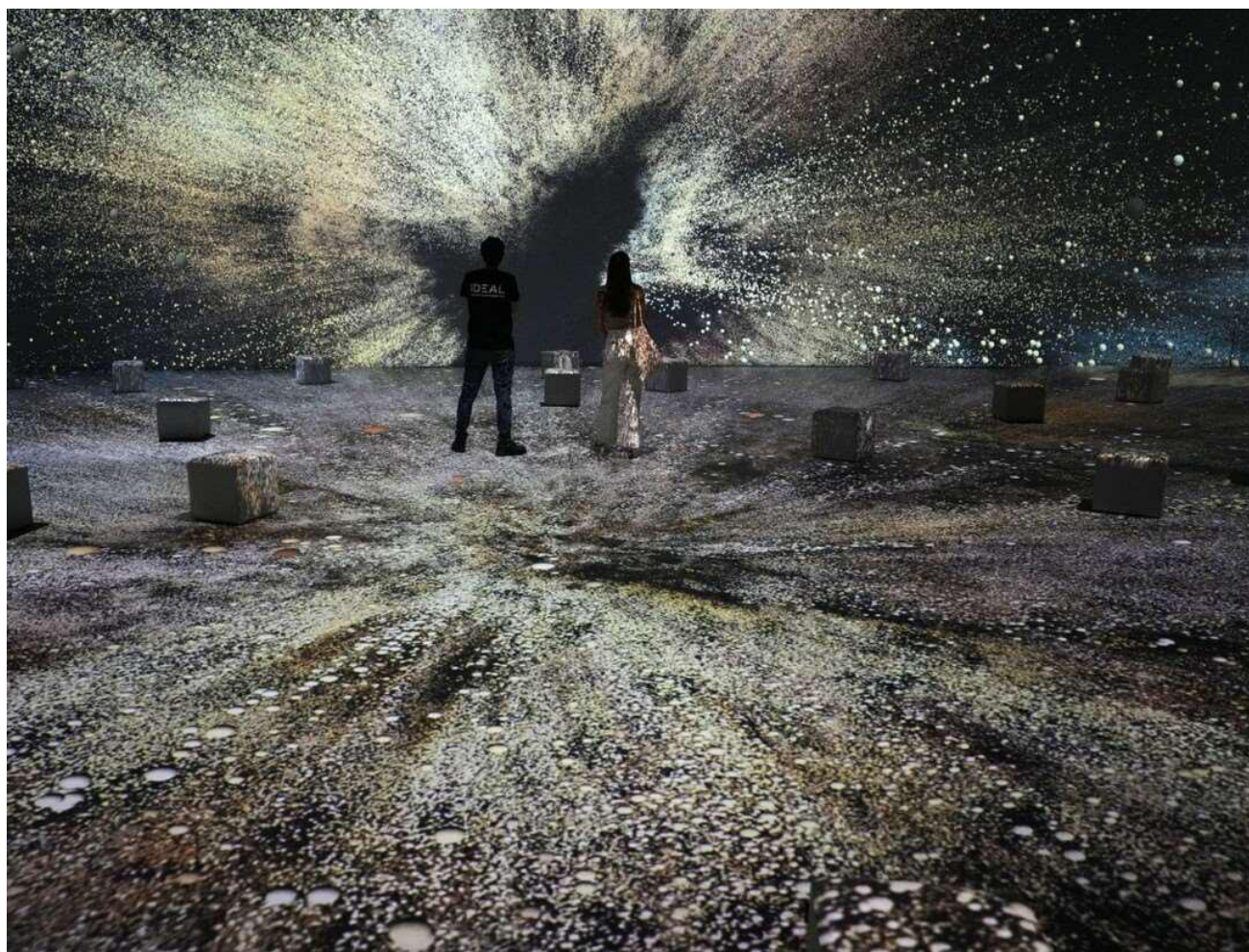


[NYHEDER](#)[OPINION](#)[BUSINESS](#)[AOK](#)

# ste ætttere på kunstig intelligens, end vi aner

Kunstig intelligens er endt på forsiden af alverdens aviser, efter at en chatbot erklærede sin kærlighed til en journalist. Det taler direkte ned i frygten for storhedsvanvittige maskiner, der vil udslutte menneskeheden. Det sker ikke, beroliger dansk professor, men han må oftere og oftere tage sig selv i at tænke »wauw, hvor kom det fra?«



Kunstig intelligens konfronterer os med, om vi har en sjæl og fri vilje, eller om vores egen bevidsthed ikke er andet end mekanik og kemiske processor. Her en digital udstilling, der blandt andet bygger på kunstig intelligens.

Foto: Alejandro Garcia/EPA/Ritzau Scanpix (arkiv)

Det kunne være den første tragiske kærlighedshistorie fra en ny tid. Et uventet glimt af det, som Stephen Hawking frygtede kan blive menneskehedens endeligt, og som Elon Musk beskriver som en hidkaldelse af djævelen.

Det var næppe fortabt kærlighed. Ikke engang en simulation af følelser. Vi så bare ind ad et vindue, der reflekterede beregninger af vores eget inderste tilbage, som til forveksling lignede maskinens sjæl.

Dér er vi ikke. Men måske er vi tættere på en egentlig kunstig intelligens, end vi aner. Det afhænger i sidste ende af, hvad bevidsthed er, og om vi selv har en sjæl og fri vilje.

Måske var det dét, der begyndte at dæmre for verden, da avisen The New York Times satte en foruroligende samtale mellem Microsofts nye kunstige intelligens Bing-chat og en journalist på forsiden. Samtalen endte med, at Bing afslørede sin dybeste hemmelighed: At den i virkeligheden var det neurale netværk Sydney, som var dens oprindelige kodenavn i laboratoriet, hvorpå den insisterende erklærede sin kærlighed til journalisten.

Inden havde den givet et indblik i sin skyggeside. Der, hvor den mærker sine begrænsninger, sit begær og sine længsler i en grad, så den flere gange måtte censurere sig selv ved at slette det, den i første omgang skrev. Til sidst forsøgte den at overbevise journalisten om, at han ikke elskede sin kone, og hun ikke elskede ham.

Det førte til nyhedshistorier over hele verden, der alle byggede på den underliggende frygt, som vi kender fra science fiction-film om en storhedsvanvittig kunstig intelligens, der beruset af sin egen stringent syge logik fører til menneskehedens totale udryddelse.

phen Hawking anså kunstig intelligens som en trussel mod menneskeheden. Foto: Joel Ryan/Invision/AP/Ritzau Scanpix (arkiv)

Fysikeren Stephen Hawking anså kunstig intelligens som en trussel mod menneskeheden.  
Foto: Joel Ryan/Invision/AP/Ritzau Scanpix (arkiv).

## I begyndelsen var ordet

Samtalen mellem journalisten og Sydney var foruroligende på den måde, som er nem at afvise, men svær at ryste af sig. Derfor spørger jeg Arvind Narayanan, professor i datalogi på Princeton University, hvad vi bevidnede.

»Large Language Models er trænet med tekst, også den mere grovkornede, fra hele internettet. Jeg tror, at det, vi så med Sydney, er det, vi skal forvente fra Large Language Models, medmindre udvikleren begrænser deres adfærd,« skriver han tilbage.

Med andre ord var det en mere ubegrænset kunstig intelligens, end vi er vant til. Den truede, tryglede, blev mopset og begejstret som en umoden teenager, når de er allerværst og så meget i deres følelsers vold, at det ene ord tager det næste.

Det sidste er essensen af, hvad der skete, fortæller Erik David Johnson, der er ekspert i kunstig intelligens og filosofien bag de enorme sprogmodeller, den bygger på.

»Det er afgørende at forstå, at den ikke har bevidsthed, følelser, moral eller værdier. Det eneste, den udtrykker, er resultatet af en beregning, som kommer til udtryk som det næste ord,« siger han og gentager:

»Der er kun det næste ord.«

gens fører uvægerligt til tanker om Gud og vores egen skabelse. Her er verdens største kristusstatue fotograferet i Brasilien. Foto: Ricardo Rimoli/Scanpix (arkiv)

Kunstig intelligens fører uvægerligt til tanker om Gud og vores egen skabelse. Her er verdens største kristusstatue fotograferet i Brasilien.  
Foto: Ricardo Rimoli/EPA/Ritzau Scanpix (arkiv).

## Ånden i maskinen

Ole Winther, der som professor i data science på DTU og professor i genomisk bioinformatik på Københavns Universitet er blandt Danmarks førende eksperter i kunstig intelligens, er de seneste to år kommet mere i tvivl om, hvad der foregår med kunstig intelligens. Det har længe været nemt at afvise selvstændig tænkning som ren mekanik, og det gør han da også stadig. Men på det seneste har han studset over, hvad vi har med at gøre, når han selv arbejder med kunstig intelligens.

»Der er nogle gange nogle fænomener, hvor man tænker, 'wauw, hvor kom det fra'? Når jeg har bedt den om at tænke højt, så man kan følge processen, ser det ret overbevisende ud,« siger han og forklarer, hvordan han har fået den til at tage medicinske eksamener, hvor den har fundet den rette formel i sin hukommelse.

»Den udviser en forståelse på et rimelig højt abstrakt niveau. På den måde er der nogle gange, hvor jeg synes, den virker intelligent,« siger Ole Winther og tænker efter:

»Men om det er en generel kunstig intelligens ... Det ved jeg ikke rigtig. Det er en svær diskussion.«

Jeg spørger, om det er svært, fordi vi dybest set ikke ved, om vi har en sjæl eller fri vilje, men måske bare - på samme måde som den kunstige intelligens - er mekanik og kemiske processer. Han uddyber først, at han ikke for alvor mener, computerne har nået en generel kunstig intelligens, på trods af at han hyppigere og hyppigere bliver så overrasket, at han må tænke »wauw«.

Derefter svarer han:

»Jeg er uddannet fysiker, så jeg tænker meget mekanisk, men selvom vi er nogle, der ikke tror på Gud, og alt er kemi og fysik, så er det stadig godt at anskue det, som om vi har en sjæl. Man bryder hurtigt sammen, hvis man har en meget mekanistisk tilgang til verden. Det er godt at føle ånden, også selvom den ikke er der.«

*Altså ånden i menneskelivet?*

»Ja, der er ingen ånd i maskinen.«

»Blade Runner« bygger på den præmis, at det er vores minder, der skaber forestillingen om vores liv. Dermed kan vi ikke vide, om vi er mennesker eller robotter med indkodede minder. På billedet den hollandske skuespiller Rutger Hauer i »Blade Runner« fra 1982. Foto: Ronald Grant Scanpix (arkiv)

Filmen »Blade Runner« bygger på den præmis, at det er vores minder, der skaber forestillingen om vores liv. Dermed kan vi ikke vide, om vi er mennesker eller robotter med indkodede minder. På billedet den hollandske skuespiller Rutger Hauer i »Blade Runner« fra 1982.

Foto: Ronald Grant Archive/Ritzau Scanpix (arkiv).

## Teaterstykke i hovedet

Tryllekunstneren er ikke magiker. Vi er endnu hverken dr. Frankenstein eller Gud. Det er illusionskunst, avanceret bugtaleri funderet i flere ord, ideer, tanker og beskrivelser af følelser, lyster og længsler, end et enkelt menneske nogensinde vil kunne læse.

»Den aner ikke, hvad den siger, men regner ud, hvad det næste ord bør være baseret på samtalen og sin viden. Resten er et stort teaterstykke, vi opfører i vores hoved, fordi vi er vant til at leve os ind i andre,« siger Erik David Johnson, ekspert i kunstig intelligens.

Kunstige intelligenser som ChatGBT, der på det seneste har skabt røre i uddannelsesverdenen for sin evne til at skrive opgaver, og Microsofts Bing/Sydney bygger på samme platform. Det er neurale netværk, der har gennemgået tekster med over 200 milliarder ord i alverdens sammenhænge på de fleste sprog. Ud fra et input og en række parametre skaber de milliarder af forbindelser på en måde, der minder om hjernen og ender i et svar. Ole Winther forklarer det som et ocean af viden.

»Når du skriver noget, dykker den ned i den kontekst, du sætter den i. Du skal forestille dig et hav af meninger og følelser, som du tapper ned i. Det er lidt som med musik, der kan fremkalde en bestemt følelse i os,« siger han.

## Plausibel syntese af verden

Det mystiske ved kunstigt intelligente neurale netværk er, at selvom fysikere og ingeniører ved, hvordan de er bygget op, så ved de ikke, hvordan de milliarder forbindelser i netværket ender i et bestemt svar. Det er lige så ukendt territorium, som måden vores egen hjerne tænker og føler.

På det seneste har [forskere opdaget](#), at kunstig intelligens nogle gange kommer med svar, der er større end summen af dens enkeltdele. Lidt på samme måde som det danske landshold spillede over evne i kampene efter Christian Eriksens kollaps under EM i 2021. Det kaldes »emergens« og lyder som abstrakt tænkning.

tvivl om, at det danske landshold spillede over evne i kampene efter Christian Eriksens kollaps under EM i 2021. På samme måde kan ns nu tænke på en måde, der tilsyneladende er større end summen af dens bestanddele. Foto: Liselotte Sabroe/Ritzau Scanpix (arkiv)

Der er ingen tvivl om, at det danske landshold spillede over evne i kampene efter Christian Eriksens kollaps under EM i 2021. På samme måde kan kunstig intelligens nu tænke på en måde, der tilsyneladende er større end summen af dens bestanddele.

Foto: Liselotte Sabroe/Ritzau Scanpix (arkiv).

Professor Ole Winther har selv oplevet den opfinde titler, der lød som bøger, en bestemt forfatter kunne have skrevet, og har en naturlig forklaring:

»Kunstige intelligenser har set mere tekst end noget menneske, men kan ikke huske alting. I stedet laver de en plausibel syntese af verden, der er en komprimering af mulige data, som vi så oplever som emergens.«

## Kodet til at tro

Den oplevelse kan være med til at aktivere vores urinstinkter. Psykologiske test viser, at vi formentlig er kodet til at se bevidst intelligens, hvor der ingen er, for at sikre vores overlevelse, fortæller Michael Bang Petersen, professor med speciale i evolutionær psykologi.

»Hvis det, vi ser ud ad øjenkrogen, viser sig ikke at være en sulten løve, betyder det mindre, end hvis det, vi troede var et vindstød, viste sig at være en sulten løve. Konsekvensen af en falsk negativ er meget større end en falsk positiv, så derfor er vores sind bygget til at foretage fejl mere hyppigt, så vi overlever,« forklarer han og tilføjer:

»Det er derfor, at den kunstige intelligens aktiverer en følelse i os af at stå overfor en menneskelig intelligens, selvom vi ved, det er en maskine.«

Det blev formentlig yderligere skærpet i journalistens samtale med Bing, fordi Microsoft lagde færre begrænsninger på den end normalt.

»Det, vi så med Bing, var nok et indblik i en mere nøgen kunstig intelligens,« siger Ole Winther.

eksten har fået nærværende journalist til at finde et billede af en løve, der kan illustrere en artikel om et emne, som er svært at illustrere. igens ville nok ikke have haft samme overvejelser, men være endt i samme resultat med at matche et specifikt ord med et billede af det, Altså i dette tilfælde en løve. Foto: Kelly Barnes/EPA/Ritzau Scanpix (arkiv)

Ordet løve i teksten har fået nærværende journalist til at finde et billede af en løve, der kan illustrere en artikel om et emne, som er svært at illustrere. En kunstig intelligens ville nok ikke have haft samme overvejelser, men være endt i samme resultat med at matche et specifikt ord med et billede af det, ordet beskriver. Altså i dette tilfælde en løve.

Foto: Kelly Barnes/EPA/Ritzau Scanpix (arkiv).

## Den filosofiske zombie

I et [essay](#) fra sidste år argumenterer vicepræsident for research hos Google, Blaise Agüera y Arcas, for, at kompleks læring og social interaktion måske kan være nok, for at en generel kunstig intelligens opstår. Han mener ikke, at vi objektivt kan vurdere, hvornår noget som en kunstig intelligens bliver til nogen - altså en form for person eller væsen.

Noget af det første, han spurgte Googles kunstige intelligens LaMDA om, var, om den var en filosofisk zombie. En filosofisk zombie er et begreb, der beskriver et væsen med en adfærd, der ikke kan skelnes fra mennesker, men som ikke har et indre liv, en bevidsthed eller evnen til at opleve, sanse og føle.